AD_____

GRANT NUMBER DAMD17-96-1-6234

TITLE: Role of Ser-Arg Proteins in the Regulation of RNA Processing

PRINCIPAL INVESTIGATOR: Benjamin J. Blencowe, Ph.D.

CONTRACTING ORGANIZATION: Massachusetts Institute of Technology
Cambridge, Masschusetts 02139

REPORT DATE: September 1998

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

DTIC QUALITY INSPECTED 3

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1998	Annual (1 Sep 97 - 31 Aug 98)

4. TITLE AND SUBTITLE
Role of Ser-Arg Proteins in the Regulation of RNA Processing

5. FUNDING NUMBERS
DAMD17-96-1-6234

6. AUTHOR(S)
Benjamin J. Blencowe, Ph.D

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

19981231 049

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

Deregulation of alternative splicing has been linked to malignant transformation in breast cancers. Therefore, to fully understand breast cancer, it will be important to identify and characterize factors that regulate the splicing process. We have previously identified a complex of nuclear matrix proteins related to Serine/Arginine-repeat (SR) splicing factors which is required for the splicing of specific pre-mRNAs (Blencowe et al., 1998). This complex contains the SR-matrix proteins of 160kDa and 300kDa (SRm160 and SRm300) which, unlike proteins of SR family of splicing factors, lack RNA Recognition Motifs (RRMs). This complex associates with pre-mRNA through multiple interactions involving SR family proteins and small nuclear ribonucleoprotein particles (snRNPs). During the past year, we have completed the sequencing of SRm300 and have investigated the function of the SRm160/300 complex in regulated splicing. We have demonstrated that SRm160/300 is required for a purine-rich exonic splicing enhancer (ESE) to promote the splicing of a pre-mRNA containing a weak 3' splice site. This function of SRm160/300 depends on the formation of an early splicing complex containing U1 snRNP and also involves interactions between SRm160/300, U2 snRNP and the ESE-binding protein Transformer2. The results thus provide evidence for a critical role for SRm160/300 as a "coactivator" in the regulation of splicing by purine-rich ESEs.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	32
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

N/A Where copyrighted material is quoted, permission has been obtained to use such material.

N/A Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

N/A Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

BB In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

BB In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

BB In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

9/24/98

PI - Signature Date

CONTENTS **PAGE**

Introduction

Deregulation of alternative splicing as been linked to malignant transformation and the formation of metastases in breast cancers. For example, specific alternative spliced forms of the cell surface adhesion glycoprotein CD44 have been correlated with invasive tumor formation. Moreover, it has been demonstrated that expression of specific alternative spliced forms of CD44 mRNAs in non-metastatic cell lines result in transition to full metastatic potential. We are interested in understanding the mechanisms underlying the regulation of alternative splicing with the long term goal of identifying and targeting trans-acting splicing factors that are involved in metastatic transitions in breast and other types of cancer.

Approximately one third of mammalian pre-mRNAs are alternatively spliced to generate protein products with functionally distinct properties (reviewed in: Smith et al., 1989; Green, 1991; Wang and Manley, 1997). Alternative splicing is a critical step in many cell differentiation and developmental pathways and must be precisely regulated. Despite the wealth of information that has emerged on the function of basal components of the spliceosome, little is known about the factors and mechanisms underlying the regulation of splice site selection.

The major spliceosome is a ~60S complex composed of four small nuclear ribonucleoprotein particles (snRNPs) and many non-snRNP splicing factors (Moore et al., 1993; Krämer, 1996). A large number of non-snRNP splicing factors have been identified which contain domains rich in serine-arginine repeats (SR proteins) (reviewed in: Fu, 1995; Manley and Tacke, 1996; Valcarcel and Green, 1996). A subgroup of these proteins, the "SR family", contain one or two N-terminal RNA recognition motifs (RRMs) and a C-terminal domain rich in serine and arginine residues (RS domain) in which many of the serines are phosphorylated. SR family proteins function at multiple stages of spliceosome assembly. They promote the formation of splicing "commitment" complexes containing U1 snRNP bound to the 5′ splice site and the U2 snRNP-auxiliary-factor (U2AF-65/35kDa) bound to the polypyrimidine tract adjacent to the 3′ splice site (Fu, 1993; Wu and Maniatis, 1993; Kohtz et al., 1994; Staknis and Reed, 1994). They also promote the subsequent formation of pre-splicing complexes containing U2 snRNP and the recruitment of U4/5/6 tri-snRNP to form assembled spliceosomes (Crispino et al., 1994; Tarn and Steitz, 1994; Roscigno and Garcia-Blanco, 1995). At each of these stages of spliceosome assembly, it is thought that SR family proteins function by promoting interactions with each other and with snRNP-associated proteins that contain RS domains (Wu and Maniatis, 1993; Kohtz et al., 1994; Fetzer et al., 1997). For example, it has been proposed that splice site recognition and pairing across introns is promoted by a network of interactions involving the association of the SR family proteins SC35 and ASF/SF2 with the U1 snRNP-70kDa protein at the 5′ splice site and with U2AF-35kDa bound to the polypyrimidine tract (Wu and Maniatis, 1993); both of the latter proteins contain short RS domains. The phosphorylated RS domains of these proteins are most likely required for the protein-protein interactions proposed to be involved in this network (Wu and Maniatis, 1993; Amrein et al., 1994; Kohtz et al., 1994; Xiao and Manley, 1997).

In addition to their roles in constitutive splicing, SR family proteins also function in the regulation of splice site selection (reviewed in: Chabot, 1996; Fu, 1995; Valcarcel and Green, 1996; Manley and Tacke, 1996). Elevated concentrations of SR family proteins promote the selection of alternative splice sites in vitro (Ge and Manley, 1990; Krainer et al., 1990; Fu et al., 1992; Zahler et al., 1993), and in vivo (Cáceres et al., 1994; Screaton et al., 1995; Wang and Manley, 1997). SR family proteins, and other RS domain-containing proteins, also function in the regulation of alternative splicing by interacting with specific intron or exon sequences called "enhancers". In a prototypic example, regulation of

alternative splicing of the *Drosophila doublesex* (*dsx*) pre-mRNA, which is part of a cascade of regulatory splicing events that determines the sex of *Drosophila*, involves the assembly of a multi-SR protein complex on an exonic splicing enhancer (ESE) within exon 4 of the *dsx* pre-mRNA (Inoue et al., 1992; Tian and Maniatis, 1993; Tian and Maniatis, 1994). The assembly of this complex, which contains the RS domain proteins Tra, Tra2, and SR family proteins, promotes the recognition of a weak, upstream, female-specific 3′ splice site, thereby promoting exon 4 inclusion. The *dsx* ESE consists of six, tandem, 13-nucleotide repeat sequences and a purine-rich element, both of which are required for efficient use of the female-specific splice site (Lynch and Maniatis, 1995). The *dsx* ESE can function in heterologous pre-mRNAs and, similarly, it can be replaced functionally by purine-rich ESEs from other alternatively spliced pre-mRNAs (Tian and Maniatis, 1992; Yeakley et al., 1996). Recently, it was demonstrated that hTra2α and hTra2β, the human homologs of *Drosophila* Tra2, preferentially bind to purine-rich ESEs containing GAA repeats and, in conjunction with specific SR family proteins, promote ESE-dependent splicing (Tacke et al., 1998). However, the mechanism by which ESEs promote splice site recognition and splicing by communicating with the general splicing machinery is not well understood.

In recent studies supported by the Breast Cancer Research Program, we have isolated and characterized the function in pre-mRNA splicing of two novel nuclear matrix proteins, the SR-matrix proteins of 160kDa and 300kD (SRm160 and SRm300; Blencowe et al., 1994, 1995, 1998). Both are novel proteins that contain multiple SR repeats but, unlike members of the SR-family of splicing factors, lack RRMs (Blencowe et al., 1998; Eldridge et al., submitted). SRm160 and SRm300 form a complex (SRm160/300) that associates with a subset of the SR-family proteins. SRm160/300 binds to pre-mRNA and promotes splicing activity through co-operative interactions with SR proteins. Interestingly however, SRm160/300 is only required for splicing of specific pre-mRNAs. In this year's report, new experiments are described which provide evidence for a critical role of SRm160/300 in the regulation of pre-mRNA splicing mediated by exonic splicing enhancer sequences (ESEs). SRm160/300 is required for a purine-rich ESE to promote splicing of a pre-mRNA containing a weak 3′ splice site. The association of SRm160/300 with this pre-mRNA requires both U1 snRNP and factors bound to the ESE. The detection of specific interactions between SRm160/300, U2 snRNP, and the ESE-binding SR repeat protein, hTra 2β, suggests that SRm160/300 functions as a coactivator of ESE-dependent splicing by mediating interactions between multiple components bound to the pre-mRNA.

These studies demonstrate a role for SRm160/300 in ESE-dependent splicing and provided a new model for how specific sets of exons are recognized and paired during regulated splicing. These studies thus provide an entrée into an investigation of the role of specific nuclear matrix proteins in the regulation of pre-mRNA processing in vitro, and in vivo. The potential role of the SRm160/300 complex in altered splicing patterns linked to tumor metastasis will be investigated.

BODY

In last year's report, the cloning of cDNAs encoding SRm160 and a functional characterization of the SRm160/300 complex in constitutive splicing was described. In this year's report, the cloning of cDNAs encoding SRm300, and the role of the SRm160/300 complex in ESE-dependent splicing, is described. A discussion of how the results of the past year address the Statement of Work (SoW) for year 2 of the CDA and a set of data figures follows the body of the report.

The 300kDa nuclear matrix antigen recognized by mAb-B4A11 is a novel SR protein, "SRm300", which lacks an RRM

The B4A11 nuclear matrix antigen was purified as a 300kDa protein from HeLa nuclear extract as described in a previous study (see Figure 4 in Blencowe et al., 1995) and was subjected to partial proteolysis with lysC. Microsequences obtained from two of the released peptides were used to search sequence databases resulting in the identification of a human expressed sequence tag cDNA (EST 186680) with a predicted ORF containing an exact sequence match to one of the peptides. The insert of EST 186680 was used as a probe to screen a phage λ cDNA library prepared from human U937 cell mRNA (kind gift of J. Borrow), from which several overlapping clones were identified. A fragment from the 5′ end of the longest cDNA clone (B4A11.7) was used as a probe to isolate additional cDNA clones extending in the 5′ direction. Multiple independent cDNA clones from both rounds of library screening were sequenced on both strands to determine the entire ORF sequence shown in Figure 1A. The corresponding cDNA sequence has been deposited in the GenBank database (accession number#). cDNA clones extending furthest in the 5′ direction contained an in-frame ORF methionine codon preceded by stop codons in all three upstream reading frames. The 3′ end of the ORF terminates in a stop codon followed by a predicted 3′ UTR of 646 bases containing a consensus polyadenylation signal and a 3′ poly A tail. The predicted ORF contained both peptide sequences obtained from the purified protein (underlined in Figure 1A). These and other data described below indicate that the predicted ORF in Figure 1A corresponds to the 300kDa B4A11 nuclear matrix antigen. Moreover, similar to other phosphorylated S/R-rich proteins, SRm300 migrates in an SDS polyacrylamide gel at a molecular weight significantly higher than its predicted mass of 246 kDa.

The predicted ORF is notable for its remarkable content of serine (S), arginine (R) and proline (P) resides (23.5%, 17.1% and 11.8%, respectively). In particular, it contains numerous clusters of SR dipeptides and two polyserine domains of unprecedented length (25 and 41 residues, respectively). Both of these sequence features are similar to features of SRm160, which also is S/R/P-rich and contains multiple SR repeats and polyserine stretches. Also similar to SRm160, SRm300 lacks an RNA Recognition Motif (RRM) typical of smaller SR proteins belonging to the SR family of splicing factors.

In addition to SR repeats and polyserine domains, SRm300 also contains three distinct types of consensus repeat sequences (Table 1). These consensus sequences are rich in S, R, P and glycine (G) or glutamic acid (E) residues and are distinct from the S, R, P-rich repeats present in SRm160. The repeat consensus sequences also do not resemble repeats in other SR-related proteins that have recently been identified (Yuryev et al., 1996; Zhang and Wu, 1998). However, like the repeats in SRm160, the SRm300 repeat consensus sequences contain potential phosphorylation sites for multiple kinases; in particular, the presence of multiple SR repeats is a strong indicator that SRm300 is a substrate for kinases such as SRPK1 and Clk1, which phosphorylate serine residues in SR repeats (Gui et al., 1994; Colwill et al., 1996). Consistent with this prediction, SRm300 is detected by the

monoclonal antibody mAb104 which detects phosphoserine residues within SR repeats, and is a substrate for SRPK1 in vitro (Blencowe et al., 1995; B.J.B. and P.A.S., unpublished observations).

The ORF in Figure 1A, with the repetitive sequences masked, was used to search the databases for related proteins. Although no previously characterized proteins were identified, the ORF amino acids 919-1479 are identical to an ORF within a previously characterized cDNA from brain (Nagase et al., 1997), and the N-terminal 167 amino acids of the ORF showed similarity to hypothetical proteins from yeast to man (Figure 1B). The percent identities of these sequences to the human protein are: *D. melanogaster* 45.5%, *C. elegans* 38.7% and *S. cerevisiae* 22.2%. It should be noted that the predicted ORF identified in the *S. cerevisiae* genomic database is only 6 amino acid residues longer than the sequence shown in Figure 1B, whereas the *D. melanogaster* and *C. elegans* ORFs C-terminal to the homologous region have not been characterized. These data indicate that the N-terminal region of SRm300 is highly conserved in metazoans and is related to a domain within a considerably smaller protein in yeast.

Association of SRm300 with splicing complexes

To verify that the predicted ORF in Figure 1A corresponds to the B4A11 nuclear matrix antigen of 300 kDa previously detected in association with SRm160, splicing complexes and nuclear structures referred to as "speckles", a polyclonal antiserum (rAb-SRm300) was raised to a GST-fusion protein containing SRm300 ORF amino acids 4 to 138. The serum was affinity purified over the GST-SRm300 fusion protein and then used in immunoblotting, immunoprecipitation and immunofluorescence microscopy experiments. The affinity purified serum specifically recognized a single antigen of ~300kDa in total HeLa cell nuclear extract (Figure 2A, lane 1). Moreover, it specifically immunoprecipitated splicing complexes from splicing reactions in an identical manner to mAb-B1C8 (which is specific for SRm160), preferentially enriching for exon-containing complexes compared to the lariat product-containing complex (Figure 2B, lane 5; compare lanes 1, 3 and 5). The corresponding pre-immune serum did not detect any nuclear protein (Figure 2A, lane 2), nor did it immunoprecipitate splicing complexes (Figure 2B, lane 4). It was also observed that rAb-SRm300, but not the pre-immune serum, specifically immunolabeled speckles in interphase nuclei in a pattern essentially identical to that of mAb-B1C8 (B.J.B, E.R. and P.A.S., unpublished observations). These properties of rAb-SRm300 confirm that the ORF in Figure 1A corresponds to the 300kDa nuclear matrix antigen recognized by mAb-B4A11, which associates with SRm160, splicing complexes, and nuclear speckles.

SRm160/300 is required for exonic splicing enhancer (ESE) function

In previous studies, it was demonstrated that the SRm160/300 complex is required for the splicing of specific pre-mRNA substrates and that it associates with pre-mRNA through multiple interactions involving SR-family proteins, U1 and U2 snRNPs (Blencowe et al., 1998). The pre-mRNAs used in these previous studies all contained relatively strong splice sites and were efficiently spliced in vitro, indicating that the SRm160/300 complex is required for processing of a subset of constitutively spliced pre-mRNA substrates. To determine if SRm160/300 is also required for exonic splicing enhancer (ESE)-dependent splicing of a pre-mRNA containing a weak 3′ splice site, pre-mRNAs containing sequences from exons 3 and 4 of the *Drosophila doublesex* (*dsx*) gene, with or without an ESE consisting of 3 or 6xGAA repeats in exon 4 (dsxΔE, dsx$(GAA)_3$ and dsx$(GAA)_6$ pre-mRNAs; kind gift of J. Yeakley and X.-D. Fu; Yeakley et al., 1996), were tested for activity in HeLa nuclear extracts in the presence or absence of SRm160/300 (Figure 3). Nuclear extracts were specifically immunodepleted of SRm160/300 with rAb-SRm160, or mock-depleted with the corresponding pre-immune serum. These extracts contained normal

levels of SR family proteins as detected by mAb104 and are identical to the preparations characterized in Blencowe et al., 1998 (see Figure 3 in this previous study).

Consistent with previous reports, splicing of the dsx pre-mRNA was dependent on GAA repeats in reactions containing the mock-depleted nuclear extract. Increasing the number of GAA repeats from 0 to 6 resulted in a significant stimulation of splicing in the mock-depleted extract (Figure 3; compare lanes 2, 4 and 6). However, immunodepletion of the SRm160/300 complex markedly reduced this level of ESE-dependent splicing (compare lanes 5 and 6). This loss of splicing activity was not due to a non-specific effect since splicing activity in the SRm160/300 depleted extract can be restored by addition of purified SRm160/300 components (Blencowe et al., 1998; data not shown). Moreover, a longer exposure of the gel shown in Figure 3 revealed that depletion of SRm160/300 did not significantly inhibit the low levels of splicing observed for the dsx$(GAA)_3$ and dsxΔE pre-mRNAs (compare lanes 3 and 4; data not shown). These results indicate that the ESE-dependent splicing of the dsx pre-mRNA requires SRm160/300.

The ESE promotes the recruitment of SRm160/300 to the dsx pre-mRNA

To investigate the mechanism by which SRm160/300 promotes the ESE-dependent splicing of the dsx pre-mRNA, it was next determined if the ESE is required for the recruitment of SRm160/300 to the pre-mRNA (Figure 4). Immunoprecipitations were performed with rAb-SRm300 and mAb-B1C8 from splicing reactions incubated with the dsxΔE, dsx$(GAA)_3$ or dsx$(GAA)_6$ pre-mRNAs. In the absence of an ESE, both antibodies immunoprecipitated low levels of pre-mRNA (lanes 8 and 11), whereas the levels of pre-mRNA (and exon-product RNA) immunoprecipitated by both antibodies increased as the numbers of GAA repeats increased (lanes 10 and 13). In several repeat experiments, the level of the dsxΔE pre-mRNA recovered was occasionally higher than that observed in Figure 4 but was always low compared to the level of recovery of the dsx$(GAA)_6$ (eg. Figure 5 and data not shown). These results indicate that the GAA-repeat ESE is important for the stable recruitment of SRm160/300 to the dsx pre-mRNA.

U1 snRNP and the ESE function together to recruit SRm160/300 and U2 snRNP to the dsx pre-mRNA

The association of SRm160/300 with the dsx pre-mRNA could be mediated by factors bound directly to the ESE, or the association may be more indirect through the formation of one or more snRNP-containing splicing complexes promoted by the ESE. To differentiate between these possibilities, the association of SRm160/300 with the dsx pre-mRNAs was assayed in splicing reactions depleted of U1 or U2 snRNPs (Figure 5).

Antibodies to SRm160 (mAb-B1C8) and SRm300 (rAb-SRm300) did not immunoprecipitate the dsx pre-mRNA, with or without an ESE, in the absence of U1 snRNP (Figure 5, lanes 17 and 20). In the absence of U2 snRNP, both antibodies immunoprecipitated increasing levels of pre-mRNA as the number of GAA-repeats increased (compare lanes 24-26 and 27-29). However, at 6xGAA repeats the level of pre-mRNA immunoprecipitation in the U2 snRNP depleted reaction was not as high as the levels observed in a mock depleted splicing reaction containing both U1 and U2 snRNPs (compare lanes 26 and 29 with lanes 8 and 11). Immunoprecipitation could be fully restored by mixing of the U1 and U2-depleted extracts indicating that the reduced levels observed in the snRNP-depleted extracts was not due to a non-specific effect (see Figure 6A, lane 12; data not shown). These results indicate that both U1 snRNP and factors bound to the ESE are required to promote the stable association of SRm160/300 with the pre-mRNA, whereas U2 snRNP is not absolutely required but further stabilizes the SRm160/300 association.

To distinguish whether binding of U1 snRNP to the dsx pre-mRNA and factors to the ESE are independent or interdependent events required for the recruitment of SRm160/300, the ability of U1 snRNP to bind to the three dsx pre-mRNAs was analyzed using a splicing complex affinity-selection assay (Figure 6A; Ryder et al., 1990). Complexes assembled on biotinylated dsx pre-mRNAs in the different snRNP-depleted nuclear extracts used in Figure 5 were selected on streptavidin agarose beads, eluted, and analyzed for their snRNP composition by northern hybridization using snRNA-specific riboprobes (refer to Materials and Methods).

U1 snRNP binds efficiently and at an approximately equivalent level to all three dsx pre-mRNAs in the mock-depleted and U2 snRNP-depleted reactions (Figure 6A, compare lanes 3-5 and 9-11). This demonstrates that binding of U1 snRNP to the dsx pre-mRNA occurs independently of the presence of an ESE. Therefore, binding of U1 snRNP, although required, is not sufficient for the recruitment of SRm160/300 to the dsx pre-mRNA; binding of factors to the ESE is also required. Moreover, consistent with the increased levels of splicing promoted by the ESE in the mock depleted extract, binding of U2, U4/U6 and U5 snRNPs to the pre-mRNA is promoted by increasing numbers of GAA repeats only in this extract (compare lanes 3-5 with 6-11).

An association between SRm160/300 and U2 snRNP

Surprisingly, depletion of U1 snRNP not only prevented the association of SRm160/300 but also the binding of U2 snRNP to the dsx pre-mRNA, even in the presence of 6xGAA repeats (Figure 6A, lanes 6-8). This indicates that U1 snRNP and ESE-bound components cooperate to recruit both SRm160/300 and U2 snRNP to the dsx pre-mRNA. By contrast, as described above, depletion of U2 snRNP resulted in a partial loss of the SRm160/300 association with the GAA repeat-containing pre-mRNAs (Figure 5, lanes 24-29). Similarly, depletion of SRm160/300 resulted in a partial loss of binding of U2 snRNP to the 6xGAA repeat dsx pre-mRNA (data not shown).

In light of this mutual interdependence for stable binding to pre-mRNA, and the parallel entry of SRm160/300 and U2 snRNP into splicing complexes (Figures 5 and 6), it was next determined if SRm160/300 and U2 snRNP interact with one another. rAb-SRm160-immunoprecipitates prepared from HeLa nuclear extract were probed for spliceosomal snRNAs (Figure 6B). rAb-SRm160, but not a corresponding pre-immune serum, specifically immunoprecipitated a subpopulation of U2 snRNP, but not the other spliceosomal snRNAs, from nuclear extract (compare lanes 2 and 3). This immunoprecipitation was not prevented by masking of U2 snRNA with an antisense oligonucleotide complementary to the branch site-pairing region (data not shown). These results indicate that U2 snRNP and SRm160/300 associate, consistent with their parallel entry into splicing complexes containing U1 snRNP and ESE-bound factors.

SRm160/300 interacts with the ESE-binding protein hTra2β

In a recent study it was demonstrated that two human homologs of the *Drosophila* alternative splicing regulator Transformer-2, hTra2α and hTra2β, bind to ESEs containing GAA repeats (Tacke et al., 1998). Unlike SR family proteins, the hTra2 proteins contain a single RRM located between two RS domains. The hTra2 proteins are present in HeLa nuclear extracts and are detected as a ~40 kDa species in SDS polyacrylamide gels by mAb104 (Tacke et al., 1998). It has been observed that rAb-SRm160 and mAb-B1C8 specifically co-immunoprecipitate a subset of SR proteins, including a 75 kDa and 40 kDa species detected by mAb104 (Blencowe et al., 1998; Y. L. and B.J.B., unpublished

observations). To determine if this 40kDa species contains one of the hTra2 proteins, immunoprecipitates from HeLa nuclear extract were probed with an affinity-purified anti-peptide antibody specific for hTra2β (kind gift of R. Tacke and J. Manley; Figure 7). mAb-B1C8, but not a control monoclonal antibody, specifically immunoprecipitated hTra2β (compare lanes 3 and 4). This coimmunoprecipitation of hTra2β by mAb-B1C8 was resistant to extensive pre-incubation of the nuclear extract with RNases (compare lanes 4 and 5; refer to Materials and Methods), indicating that the association between hTra2β and SRm160/300 is mediated by protein-protein interactions and is not "tethered" by endogenous RNA in the nuclear extract. Probing of the same immunoblot with mAb104 revealed a comparable level of enrichment for the mAb104-reactive 40 kDa antigen(s) indicating that a significant fraction of the 40 kDa protein(s) immunoprecipitated by mAb-B1C8 may correspond to hTra2β (data not shown). Similar co-immunoprecipitation results were obtained using rAb-SRm160 (data not shown). We conclude that SRm160/300 interacts with the ESE-binding protein hTra2β and that this interaction, in conjunction with additional interactions involving other SR proteins, U1, and U2 snRNPs, is critical for the promotion of splicing by a GAA-repeat ESE.

Discussion and Conclusions

In research performed during the past year (Year 2 of the CDA) it was demonstrated that the 300kDa subunit of the SRm160/300 splicing "coactivator" is a novel protein with a remarkably high content of serine (S), arginine (R) and proline (P) residues and is named "SRm300" (the SR- nuclear matrix protein of 300kDa). Similar to SRm160, SRm300 contains multiple clusters of SR dipeptide repeats but lacks an RNA recognition motif (RRM) found in the SR family of splicing factors. SRm300 also contains unusual sequence features including many copies of three different S/R/P-rich motifs and two long polyserine domains. It is demonstrated that SRm160/300 is required for a GAA-repeat exonic splicing enhancer (ESE) to promote the splicing of a pre-mRNA derived from the *Drosophila double sex* (*dsx*) gene, which contains a weak 3′ splice site. The association of SRm160/300 with this pre-mRNA requires the formation of a complex containing both U1 snRNP and factors bound to the ESE. The detection of specific interactions between SRm160/300, U2 snRNP, and the ESE-binding SR domain protein, hTra2β, indicates that SRm160/300 may promote ESE-dependent splicing by mediating critical interactions between U1 snRNP bound to the 5′ splice site, hTra2β bound to the ESE, and U2 snRNP bound to the pre-mRNA branch site. These results support a model in which SRm160/300 functions as a coactivator of ESE-dependent splicing by bridging between SR protein "activators" bound to an ESE and "basal" snRNP components of the spliceosome (Figure 8).

The results indicate the critical importance of multiple cooperative interactions involving U1 snRNP and SR proteins in the recruitment of SRm160/300 to an ESE-dependent pre-mRNA; the association of SRm160/300 with the dsx pre-mRNA in the present study was prevented by the depletion of U1 snRNP and was weak in the absence of the ESE. U2 snRNP has similar factor requirements for recruitment. Thus U1 snRNP, in addition to promoting the stable binding of U2 snRNP to constitutively spliced pre-mRNAs (Barabino et al., 1990), is also required for promoting the stable binding of U2 snRNP to an ESE-dependent substrate. SRm160/300 does not detectably interact with U1 snRNP in the absence of pre-mRNA, or with U1 snRNP bound to dsx pre-mRNA lacking an ESE. Therefore, it is likely that one or more factors that associate with pre-mRNA after the binding of U1 snRNP to the 5′ splice site and factors to the ESE, recruit both SRm160/300 and U2 snRNP. The specific interaction detected between SRm160/300 and U2 snRNP is consistent with the parallel entry of these components into splicing complexes. The observation that depletion of SRm160/300 or U2 snRNP weakens but does not prevent the association of the other component with the pre-mRNA (this study; A.E., P.A.S. and B.J.B., unpublished observations), provides evidence that these components can also interact with splicing complexes independently. The interactions between SRm160/300, U2 snRNP and ESE-bound components detected in this study may occur in conjunction with previously proposed interactions required for ESE function, for example, the interaction between specific SR family proteins such as ASF/SF2 which bind to purine-rich ESEs, and U2AF-35kDa which binds to the polypyrimidine tract through U2AF-65kDa (see introduction). However, unlike SRm160/300, binding of U2AF to the pre-mRNA is not known to require U1 snRNP. This fundamental difference suggests that SRm160/300 may play a more direct role than U2AF in promoting cross-intron interactions on ESE-dependent substrates.

The requirement of SRm160/300 for the ESE-dependent recognition of a weak 3′ splice site is consistent with it having a critical role in the regulation of splice site selection. SRm160/300 may promote the joining of specific pairs of exons dependent on GAA-repeat enhancer sequences. However, the requirement for SRm160/300 for splicing also depends

on other sequence features in a pre-mRNA. For example, we have recently observed that SRm160/300 is required for splicing in vitro of a subset of pre-mRNAs that contain relatively strong splice sites and are not dependent on a GAA-repeat ESE (Blencowe et al., 1998). Moreover, it has also been observed that SRm160/300 is not required for increased splicing activity promoted by a GAA-repeat ESE in the 3´ exon of a two exon β-globin pre-mRNA substrate (A.E., P.A.S. and B.J.B., unpublished observations). However, this substrate contains a relatively strong 3´ splice site and is spliced in an SRm160/300 depleted reaction in the absence of the ESE. Thus, a requirement for SRm160/300 may be determined by the relative strength of a 3´ splice site and by additional sequence features in a pre-mRNA besides an ESE.

A multisubunit complex containing the *Drosophila* RS domain proteins Tra2, Tra and the SR family protein RBP1/SRp20 associates with the wild type *dsx* pre-mRNA ESE and promotes the recognition of the upstream, weak, 3´ splice site (Tian and Maniatis, 1992; Tian and Maniatis, 1993; Heinrichs and Baker, 1995; Lynch and Maniatis, 1996). Furthermore, individual SR-family proteins such as ASF/SF2 can promote the splicing of pre-mRNAs dependent on purine-rich ESEs (eg. Lavigueur et al., 1993; Sun et al., 1993; Ramchatesingh et al., 1995; Tacke and Manley, 1995; Yeakley et al., 1996). However, for some substrates tested, ASF/SF2 was not sufficient to promote ESE-dependent splicing; an additional fraction, "NF20-40", was required (Tacke et al., 1997). The GAA-repeat enhancer sequence used in the present study was shown to bind a 37kDa protein in nuclear extract which immunoreacted with mAb104 but which appeared to be distinct from SR family proteins (Yeakley et al., 1996). The likely identity of this species is one or both of the human homologs of the *Drosophila* splicing regulator Tra2, which preferentially binds to ESEs containing GAA repeats (Tacke et al., 1998). Although hTra2 proteins are present in the NF20-40 fraction and are required for GAA-repeat ESE function, purified hTra2 proteins do not functionally replace the NF20-40 fraction (Tacke et al., 1998). SRm160/300 may correspond to the missing factor, as it is detected in the NF20-40 fraction (B.J.B., unpublished observations) and cooperates with SR family proteins in promoting splicing activity (Blencowe et al., 1998). The detection of interactions between SRm160/300 and a subset of SR proteins including hTra2β further indicates that it could promote ESE-dependent splicing by forming associations with these proteins bound to a GAA-repeat ESE (this study, Blencowe et al., 1998; Y. L. and B.J.B., unpublished observations).

In summary, the SRm160/300 complex functions in the ESE-dependent splicing of a dsx pre-mRNA by forming multiple interactions with factors bound directly to the ESE and snRNPs bound at splice sites. Since the RS domains of different SR proteins are known to interact (Wu and Maniatis, 1993; Amrein et al., 1994; Kohtz et al., 1994; Xiao and Manley, 1997), the large size and SR-rich nature of the SRm160 and SRm300 proteins is consistent with their proposed role in the formation of multiple interactions with other RS domain proteins bound to pre-mRNA (Figure 8). The ratio of SRm160/300 to other specific RS domain proteins could regulate the selection of splice sites in alternatively spliced pre-mRNAs, as well as influence the activity of constitutively spliced pre-mRNAs (Blencowe et al., 1998). Moreover, the abundant consensus phosphorylation sites and observed phosphorylation of SRm160 and SRm300 indicates that the regulation of splicing by differential phosphorylation mechanisms could be influenced by kinases and phosphatases that target SRm160/300. Finally, the stable association of SRm160/300 with the non-chromatin "matrix" of the nucleus indicates that it may function in association with this substructure in vivo (Blencowe et al., 1994). In particular, it is possible that SRm160/300 functions as a substructure on which specific pairs of splice sites are juxtaposed following their initial recognition by U1 snRNP and SR proteins. It is also tempting to speculate that SRm160/300 is involved in deregulated splicing events

associated with metastatic transitions in breast and other cancers. These possibilities will be investigated in remaining years of the CDA research.

Relevance to the Statement of Work (SoW) objectives.

A revised SoW was submitted as part of a request for transfer of the CDA to the University of Toronto and is included below. In collaboration with Goran Bauren in the laboratory of Phillip Sharp at MIT, we are currently analyzing the influence of overexpression of wild type and mutant derivatives of SRm160 on nuclear organization of splicing and on splicing of pre-mRNAs in vivo. These studies will address item 6 of the SoW and will be summarized in a future report. Experiments described in the Year 1 report and the present report center on defining the function of the SRm160/300 complex and identify sequences to which it is recruited in exonic RNA. These experiments address and extend objectives outlined in the SoW items 1, 2, 5. Specific experiments, such as the cross-linking strategy outlined in Task 3 will not be performed since we have already defined by alternative routes pre-mRNA sequences and factors with which the SRm160/300 complex associates in splicing complex. To extend these studies, we will determine which ESE sequences function optimally in conjunction with SRm160/300 using a randomization-selection strategy.

REVISED STATEMENT OF WORK (August 25, 1998)

Note: *The original SoW for the CDA follows. Proposed modifications to this are outlined in italicized type. We have already completed the majority of the experiments outlined in Tasks 1, 3, 5, 6 and other experiments listed in these Tasks are in progress. Due to our unanticipated discovery of an association between the hyperphosphorylated large subunit of RNA polymerase II (pol IIo) and splicing components (refer to the Annual Report for Year 1), current experiments are also directed at investigating mechanisms underlying the coupling of transcription and splicing. Therefore, changes to Tasks 2, 4 and 7 have been proposed to allow the extension of this important and timely discovery. It should also be noted that the original experiments in Tasks 2 and 4 (which lead to Task 7) have recently been performed in the laboratory of Dr. Susan Berget (Baylor College, Texas). We therefore propose to maintain our focus on the investigation of functions of novel SR-matrix proteins in pre-mRNA processing and mechanisms underlying the coupling of transcription and splicing.*

Task 1: Months 1-12: Molecular genetic approaches and in vitro assays will be used to determine the functions of SR proteins in pre-mRNA splicing. An existing clone for the matrix SR protein, SRm160, will be used to express wild type and mutant derivatives for assaying activity in SRm160-depleted splicing extracts. Depleted extracts will be prepared using an affinity purified polyclonal antiserum (in progress). Antisera to SRm160 and SRm300 matrix antigens will be used to investigate interactions of these SR-related proteins with splicing complexes. Interactions of these factors with exon-RNAs will be mapped by an RNAse protection assay.

Task 2: Months 6-12: Concurrent with Task 1, SR-enriched fractions will be prepared from different cell types for comparative screening of SR protein compositions. Enriched fractions will be prepared from cell lines available in the MIT community, including normal and malignant breast cell lines. SR proteins showing altered levels in correlation with tumorigenesis will be tested for reactivity with panels of SR reactive-mAbs.

Instead of Task 2, as explained above, we propose to investigate interactions between the transcription and splicing processes. Dr. Susan McCracken, who is a Research Associate in the lab and is funded by the CDA, is developing in vitro assays for analyzing interactions between the transcriptional machinery and pre-mRNA processing. This will extend our

previous observation that pol IIo is associated with splicing complexes (see above, previous annual reports), and her previous experiments demonstrating that the Carboxyl Terminal Domain (CTD) of RNA polymerase influences multiple steps in pre-mRNA processing.

Task 3: Months 13-24: RNA sequence elements required for the interactions of the SR related antigens (Task 1) will be further analyzed and narrowed down. Pre-mRNAs site-specifically derivatized to contain a cross-linking reagent within the binding element(s), will be used to probe for protein components which interact in a splicing-dependent manner. Cross-linking species will be screened for reactivity with panels of available antibodies to splicing factors, including SR proteins. Unidentified proteins which cross-link and look interesting, will be purified, microsequenced, and cloned.

Task 4: Months 13-24: Proteins identified in the SR-screening in Task 2, will be further purified for obtaining microsequences. Unless previously cloned, cDNAs for these proteins will be isolated and expressed to obtain specific antisera; these antisera will be used to confirm isolation of the correct malignancy-altered factor.

Instead of Task 4, the coupled transcription-splicing assay proposed in Task 2 (in progress) will be exploited to investigate the quantitative and qualitative influence of the transcriptional machinery, and the CTD in particular, on splicing. The possible influence of promoter structure on splice site selection will be tested using different promoters in conjunction with different pre-mRNA reporters.

Task 5: Months 24-36: cDNA clones isolated for new factors identified by the cross-linking experiments in Task 2 will be expressed; recombinant proteins will be used to immunize rabbits (it is anticipated that approx 2-4 rabbits will be used per 2 year period). Reactive antisera will be used to investigate the activity of the proteins identified by cross-linking in splicing, and also interactions with other factors involved in splicing. If SRm300 is identified in the assays, we will make use of existing microsequence data to clone this SR-related protein.

Task 6: Months 36-48: Wild type and mutant derivatives of SRm160 protein will be transfected into cells to determine possible effects on pre-mRNA processing and nuclear structure. Possible effects on splicing and transport will be determined by RNAse protection. Effects on nuclear structure and organization will be investigated by confocal microscopy.

Task 7: Months 36-48: SR proteins identified and cloned in Tasks 2 and 4 will be expressed in the cell-type lacking the corresponding protein. Resulting changes in splicing patterns of specific pre-mRNAs will be assayed by RT pcr and RNAse protection methods. Changes in cell growth properties and morphology will be followed by microscopy and with available reagents to appropriate cell surface markers (eg. CD44).

Factors potentially involved in the coupling of transcription and splicing will be immunodepleted from the in vitro system described in Tasks 2 and 4 and differential effects on coupled vs. uncoupled splicing will be assayed. We are currently developing antibodies to new SR proteins implicated in coupling.

References

Abovich, N. and M. Rosbash (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**: 403-412.

Alzhanova-Ericsson, A. T., X. Sun, N. Visa, E. Kiseleva, T. Wurtz and B. Daneholt (1996). A protein of the SR family binds extensively to exonic Balbiani ring pre-mRNA and accompanies the RNA from gene to the nuclear pore. *Genes & Dev.* **10**: 2881-2893.

Amrein, H., M. L. Hedley and T. Maniatis (1994). The role of specific protein-RNA interactions and protein-protein interactions in positive and negative control of pre-mRNA splicing by *transformer 2*. *Cell* **76**: 735-746.

Barabino, S. M. L., B. J. Blencowe, U. Ryder, B. S. Sproat and A. I. Lamond (1990). Targeted snRNP depletion reveals an additional role for mammalian U1 snRNP in spliceosome assembly. *Cell* **63**: 293-302.

Berget, S. (1995). Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**: 2411-2414.

Black, D. (1995). Finding splice sites within a wilderness of RNA. *RNA* **1**: 763-771.

Blencowe, B. J., R. Issner, J. Kim, P. McCaw and P. A. Sharp (1995). New proteins related to the Ser-Arg family of splicing factors. *RNA* **1**: 852-865.

Blencowe, B. J., J. A. Nickerson, R. Issner, S. Penman and P. A. Sharp (1994). Association of nuclear matrix antigens with exon-containing splicing complexes. *J. Cell Biol.* **127**: 593-607.

Blencowe, B. J., B. S. Sproat, U. Ryder, S. Barabino and A. I. Lamond (1989). Antisense probing of the human U4/U6 snRNP with biotinylated 2'-OMe RNA oligonucleotides. *Cell* **59**: 531-539.

Bregman, D., L. Du, S. van der Zee and S. Warren (1995). Transcription-dependent redistribution of the large subunit of RNA polymerase II to discrete nuclear domains. *J Cell Biol* **129**: 287-298.

Brendel, V., P. Bucher, I. Nourbakhsh, B. E. Blaisdell and S. Karlin (1992). Methods and algorithms for statistical analysis of protein sequences. *Proc. Natl. Acad. Sci. USA.* **89**: 2002-2006.

Caceres, J. F., T. Mistelli, G. R. Screaton, D. L. Spector and A. R. Krainer (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J. Cell Biol.* **138**: 225-238.

Carter, K. C., K. L. Taneja and J. B. Lawrence (1991). Discrete nuclear domains of poly(A) RNA and their relationship to the functional organization of the nucleus. *J. Cell Biol.* **115**: 1191-1202.

Chabot, B. (1996). Directing alternative splicing: cast and scenarios. *Trends in Genetics* **12**: 472-478.

Colwill, K., T. Pawson, B. Andrews, J. Prasad, J. Manley, J. Bell and P. Duncan (1996). The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J* **15**: 265-275.

Crispino, J. D., B. J. Blencowe and P. A. Sharp (1994). Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science* **265**: 1866-1869.

Crispino, J. D., J. E. Mermoud, A. I. Lamond and P. A. Sharp (1996). Cis-acting elements distinct from the 5' splice site promote U1-independent pre-mRNA splicing. *RNA* **2**: 664-673.

Du, L. and S. Warren (1997). A functional interaction between the carboxy-terminal domain of RNA polymerase II and pre-mRNA splicing. *J Cell Biol* **136**: 5-18.

Fakan, S. and E. Puvion (1980). The ultrastructural visulazation of nuclear and extranucleolar RNA synthesis and distribution. *Int. Rev. Cyt.* **65**: 255.

Fay, F., K. Taneja, S. Shenoy, L. Lifshitz and R. Singer (1997). Quantitative digital analysis of diffuse and concentrated nuclear distributions of nascent transcripts, SC35 and poly(A). *Exp Cell Res* **231**: 27-37.

Fu, X.-D. (1995). The superfamily of arginine/serine-rich splicing factors. *RNA* **1**: 663-680.

Fu, X.-D. and T. Maniatis (1990). Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **343**: 437-441.

Gui, J.-F., W. S. Lane and X.-D. Fu (1994). A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature* **369**: 678-682.

Harlow, E. and D. Lane (1988). Antibodies: a laboratory manual. Cold Spring Harbor Laboratory Press. New York.

Hedley, M., H. Amrein and T. Maniatis (1995). An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine-rich splicing factor. *Proc Natl Acad Sci U S A* **92**: 11524-11528.

Heinrichs, V. and B. S. Baker (1995). The Drosophila SR protein RBP1 contributes to the regulation of doublesex alternative splicing by recognizing RBP1 RNA target sequences. *EMBO J* **14**: 3987-4000.

Huang, S., T. Deerinck, M. Ellisman and D. Spector (1994). In vivo analysis of the stability and transport of nuclear poly(A)+ RNA. *J Cell Biol* **126**: 877-899.

Huang, S. and D. Spector (1996). Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription. *J Cell Biol* **133**: 719-732.

Jackson, D. A. and P. R. Cook (1985). Transcription occurs at the nucleoskeleton. *EMBO (Eur. Mol. Biol. Organ.) J.* **4**: 919-925.

Kim, E., L. Du, D. Bregman and S. Warren (1997). Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. *J Cell Biol* **136**: 19-28.

Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Luhrmann, M. A. Garcia-Blanco and J. L. Manley (1994). Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* **368**: 119-124.

Krainer, A. R., G. C. Conway and D. Kozak (1990). Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev.* **4**: 1158-1171.

Kramer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *1996* **65**: 367-409.

Lamond, A. I. and M. Carmo-Fonseca (1993). The coiled body. *Trends Cell Biol.* **3**: 198-204.

Lawrence, J., K. Carter and X. Xing (1993). Probing functional organization within the nucleus: is genome structure integrated with RNA metabolism? *Cold Spring Harbor Symp Quan Biol* **58**: 807-818.

Li, H. and P. M. Bingham (1991). Arginine/serine-rich domains of the *su(w^a)* and *tra* RNA processing regulators target proteins to a subnuclear compartment implicated in splicing. *Cell* **67**: 335-342.

Lynch, K. and T. Maniatis (1996). Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosphila doublesex* splicing enhancer. *Genes & Dev* **10**: 2089-2101.

Manley, J. and R. Tacke (1996). SR proteins and splicing control. *Genes & Dev* **10**: 1569-1579.

Mattern, K., R. v. Driel and L. d. Jong (1997). Composition and structure of the nuclear matrix. Nuclear Structure and Gene Expression. New York, Academic Press. 87-110.

McCracken, S., N. Fong, K. Yankulov, S. Ballantyne, G. Pan, J. Greenblatt, S. Patterson, M. Wickens and D. L. Bentley (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**: 357-361.

Meier, U. and G. Blobel (1992). Nopp140 shuttles on tracks between nucleus and cytoplasm. *Cell* **70**: 127-138.

Mermoud, J., C. Calvio and A. I. Lamond (1994a). Uncovering the role of Ser/Thr protein phosphorylation in nuclear pre-mRNA splicing. *Adv Prot Phosphatases* **8**: 99-118.

Mermoud, J., P. T. W. Cohen and A. I. Lamond (1994b). Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J.* **13**: 5679-5688.

Misteli, T., J. F. Caceres and D. L. Spector (1997). The dynamics of a pre-mRNA splicing factor in living cells. *Nature* **387**: 523-527.

Misteli, T. and D. Spector (1996). Serine/threonine phosphatase 1 modulates the subnuclear distribution of pre-mRNA splicing factors. *Mol Biol Cell* **7**: 1559-1572.

Misteli, T. and D. Spector (1997). Protein phosphorylation and the nuclear organization of pre-mRNA splicing. *Trends in Cell Biol* **7**: 135-138.

Mortillaro, M., B. Blencowe, X. Wei, H. Nakayasu, L. Du, S. Warren, P. Sharp and R. Bezerkney (1996). A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc Natl Acad Sci U S A* **93**: 8253-8257.

Nickerson, J. A., G. Krockmalnic, K. M. Wan, C. D. Turner and S. Penman (1992). A normally masked nuclear matrix antigen that appears at mitosis on cytoskeleton filaments adjoining chromosomes, centrioles and midbodies. *J. Cell Biol.* **116**: 977-987.

Penman, S., B. Blencowe and J. Nickerson (1997). The nuclear matrix: past and present. Nuclear Structure and Gene Expression. New York, Academic Press. 3-24.

Reed, R. (1996). Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr Opin Genet Dev* **6**: 215-220.

Roscigno, R. and M. Garcia-Blanco (1995). SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA* **1**: 692-706.

Roth, M. B., C. Murphy and J. G. Gall (1990). A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. *J. Cell. Biol.* **111**: 2217-2223.

Sharp, P. (1994). Split genes and RNA splicing. *Cell* **6**: 805-815.

Smith, H. C., S. G. Harris, M. Zillmann and S. M. Berget (1989). Evidence that a nuclear matrix protein participates in pre-mRNA splicing. *Exp. Cell Res.* **182**: 521-533.

Spector, D. (1993). Macromolecular domains within the cell nucleus. *Ann Rev Cell Biol* **9**: 265-315.

Spector, D. L., X.-D. Fu and T. Maniatis (1991). Associations between distinct pre-mRNA splicing components and the cell nucleus. *EMBO J.* **10**: 3467-3481.

Spector, D. L., W. H. Schrier and H. Busch (1983). Immunoelectron microscopic localization of snRNPs. *Biol. Cell.* **49**: 1-10.

Steinmetz (1997). Pre-mRNA processing and the CTD of RNA polymerase II: the tail that wags the dog? *Cell* **89**: 491-494.

Tian, M. and T. Maniatis (1992). Positive control of pre-mRNA splicing in vitro. *Science* **256**: 237-240.

Tian, M. and T. Maniatis (1993). A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA. *Cell* **74**: 105-114.

Tronchere, H., J. Wang and X.-D. Fu (1997). A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA. *Nature* **388**: 397-400.

Valcarcel, J. and M. R. Green (1996). The SR protein family: pleiotropic functions in pre-mRNA splicing. *Trends in Biochem. Sci.* **21**: 296-301.

Verheijen, R., H. Kuijpers, P. Vooijs, W. v. Venrooij and F. Ramaekers (1986). Distribution of the 70k U1 RNA-associated protein during interphase and mitosis:

correlation with other U RNP particles and proteins of the nuclear matrix. *J. Cell Sci.* **86**: 173-190.

Vincent, M., P. Lauriault, M. Dubois, S. Lavoie, O. Bensaude and B. Chabot (1996). The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. *Nucleic Acids Res* **24**: 4649-4652.

Vogelstein, B. and B. F. Hunt (1982). A subset of small nuclear ribonucleoprotein particle antigens is a component of the nuclear matrix. *Biochem. Biophys. Res. Commun.* **105**: 1224-1232.

Wan, K., J. A. Nickerson, G. Krockmalnic and S. Penman (1994). The SRm160 protein is in the dense assemblies of the nuclear matrix and relocates to the spindle and pericentriolar filaments at mitosis. *Proc. Natl. Acad. Sci. USA* **91**: 594-598.

Wansink, D. G., W. Schul, I. v. d. Kraan, B. v. Steensel, R. v. Driel and L. d. Jong (1993). Fluorescent labeling of Nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J. Cell Biol.* **122**: 283-293.

Wu, J. Y. and T. Maniatis (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**: 1061-1070.

Xiao, S. H. and J. L. Manley (1997). Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes and Dev.* **11**: 334-344.

Xing, Y., C. Johnson, P. M. Jr, J. McNeil and J. Lawrence (1995). Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains. *J Cell Biol* **131**: 1635-1647.

Xing, Y., C. V. Johnson, P. R. Dobner and J. B. Lawrence (1993). Higher level organization of individual gene transcription and splicing. *Science* **259**: 1326-1330.

Yuryev, A., M. Patturajan, Y. Litingtung, R. Joshi, C. Gentile, M. Gebara and J. Corden (1996). The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc. Natl. Acad. Sci. USA.* **14**: 6975-6980.

Zahler, A. M., W. S. Lone, J. A. Stalk and M. B. Roth (1992). SR proteins: A conserved family of pre-mRNA splicing factors. *Genes Dev.:* 837-847.

Zeitlin, S., A. Parent, S. Silverstein and A. Efstratiatdis (1987). Pre-mRNA splicing and the nuclear matrix. *Mol. Cell. Biol.* **7**: 111-120.

Zeitlin, S., R. C. Wilson and A. Efstratiatdis (1989). Autonomous splicing and complementation of in vivo assembled spliceosomes. *J. Cell Biol.* **108**: 765-777.

Year 2 Publications (1997-1998)

1. Penman, S., Blencowe, B.J., and Nickerson, J.A. (1997). The nuclear matrix: past and present. In: Nuclear Structure and Gene Expression. New York, Academic Press. p.3-24.

3. Blencowe, B.J., Issner, R., Nickerson, J.A. and Sharp, P.A. (1998) A coactivator of Pre-mRNA splicing. Genes and Dev. 12: 996-1009.

2. Blencowe, B.J. and Lamond, A.I. (1998). Purification and depletion of RNP particles by antisense affinity chromatography. Methods in Molecular Biology. In press.

4. Eldridge, A., Issner, R., Li, Y., Reifenberg, E., Sharp, P.A., and Blencowe, B.J. (1998) The SRm160/300 splicing coactivator is required for exon-enhancer function. Submitted to Genes and Dev.

Figure Legends

Figure 1. The SRm300 ORF and related sequences.
1A. Amino acid sequence of the SRm300 ORF. SR/RS dipeptides are highlighted in black boxes, polyserine domains and repeat sequences occurring ten or more times in white boxes (see Table 1). Microsequences of lys-C peptides derived from purified SRm300 are underlined. The SRm300 nucleotide sequence has been deposited in the GenBank database (accession#).

1B. Multiple alignment of sequences homologous to the N-terminal 159 amino acids of the human SRm300 ORF (Hs). Partial ORF sequences from *D. melanogaster* (Dm), *C. elegans* (Ce) and *S. cerevisiae* (Sc) were identified by BLAST searches of the expressed sequence tag (EST) and genomic sequence databases and aligned using the Clustal algorithm. The Dm sequence corresponds to a contig of two EST cDNA sequences (accession numbers: AA538757 and AA263722), whereas the Ce and Sc sequences were derived from conceptual translation of gDNA sequences (accession numbers: AL020986 and 2131522, respectively). Additional ORF sequences were identified in mouse EST cDNAs that are greater than 98% identical to the Hs sequence and a short ORF, identical to amino acids 1-18 of the Ce ORF, was identified in a B. Malayi EST cDNA (data not shown).

Figure 2. Association of SRm300 with splicing complexes
2A. Total HeLa nuclear extract was separated on a 10% SDS-polyacrylamide gel and immunoblotted with an antigen-affinity purified polyclonal antiserum raised to a GST-SRm300 fusion protein containing SRm300 amino acids 4 to 138 (rAb-SRm300, lane 1), and the corresponding pre-immune serum (lane 2).

2B. Immunoprecipitation of splicing complexes with rAb-SRm300 from reactions incubated for 40 min containing PIP85A pre-mRNA. RNA recovered following immunoprecipitation (lanes 2-5) and RNA recovered directly from a parallel splicing reaction (lane 1), was separated on a 15% denaturing polyacrylamide gel. RNA loaded in lane 1 represents 25% of the total amount recovered, whereas RNA loaded in lanes 2-5 from each immunoprecipitation represents 50% of the total amounts recovered. Immunoprecipitations were performed with a non-specific control Ab (rabbit anti-mouse, lane 2), mAb-B1C8 (lane 3), rAb pre-immune serum (lane 4) and rAb-SRm300 (lane 5).

Figure 3. The SRm160/300 complex is required for ESE-dependent splicing.
Depletion of SRm160/300 inhibits splicing of a *Drosophila doublesex* (dsx) pre-mRNA containing a GAA-repeat ESE in the 3′ exon. Splicing reactions containing SRm160/300-depleted extract (lanes 1,3,5), or nuclear extract mock depleted with pre-immune serum (lanes 2,4,6), were incubated for 1 hr with a dsx pre-mRNA with no enhancer sequence (dsxΔE; lanes 1,2), with a dsx pre-mRNA containing 3xGAA repeats ($dsx(GAA)_3$; lanes 3,4), or with a dsx pre-mRNA containing 6xGAA repeats ($dsx(GAA)_6$; lanes 5,6).

Figure 4. Efficient binding of SRm160/300 to the dsx pre-mRNA requires GAA repeats .
Splicing complexes were immunoprecipitated using mAb-B1C8 and rAb-SRm300 from reactions incubated for 40 min with the three dsx pre-mRNA substrates: dsxΔE (lanes 1,8,11), $dsx(GAA)_3$ (lanes 2,9,12), or $dsx(GAA)_6$ (lanes 3-7,10,13). RNA recovered directly from splicing reactions (lanes 1-4, 6) and RNA recovered following immunoprecipitation (lanes 5, 7-13), was separated on a 7% denaturing polyacrylamide gel. 25% of the total amount of RNA from the "Totals" and 50% of the total amount of RNA recovered from the "Pellets" was loaded. Immunoprecipitations were performed with a non-specific control Ab (rabbit anti-mouse, lane 5; the corresponding total is in lane 4),

rAb pre-immune serum (lane 7; the corresponding total is in lane 6), mAb-B1C8 (lanes 8-10) and rAb-SRm300 (lanes 11-13). The corresponding totals for the latter two sets are in lanes 1-3.

Figure 5. The association of SRm160/300 with the dsx pre-mRNA requires U1 snRNP in addition to ESE-bound factors.
Immunoprecipitations were performed using mAb-B1C8 (lanes 6-8, 15-17, 24-26) and rAb-SRm300 (lanes 9-11, 18-20, 27-29) from sets of splicing reactions incubated for 40 min containing either a control (mock-depleted) nuclear extract (lanes 1-11), a U1 snRNP-depleted nuclear extract (lanes 12-20) or a U2 snRNP-depleted nuclear extract (lanes 21-29). Each set of reactions was incubated with the three dsx pre-mRNAs described in Figure 3, as indicated above the panel. RNA recovered directly from splicing reactions (lanes 1-4, 12-14, 21-23) and RNA recovered following immunoprecipitation (lanes 5-11, 15-20, 24-29), was separated on a 7% denaturing polyacrylamide gel. The amounts of RNA loaded are as described in Figure 4. A control immunoprecipitation was performed with pre-immune serum (lane 5) from a reaction containing mock-depleted nuclear extract and the dsx$(GAA)_6$ pre-mRNA. The corresponding "total" is shown in lane 4).

Figure 6. Interactions between SRm160/300 and snRNPs in the assembly of dsx splicing complexes.
6A. U1 snRNP binds to the dsx pre-mRNA independently of the ESE, whereas the ESE and U1 snRNP are required for the assembly of U2, U4/U6 and U5 snRNPs on the dsx pre-mRNA. Biotinylated dsx pre-mRNAs were incubated in mock-depleted or snRNP-depleted splicing reactions for 40 min prior to affinity selection on streptavidin agarose. RNA recovered from the beads was separated on a 10% denaturing polyacrylamide gel and analyzed by northern hybridization using riboprobes specific for the five spliceosomal snRNAs. Lanes 3-11 contain RNA recovered after affinity selection with biotinylated dsx pre-mRNAs and lane 2 contains RNA recovered after a control selection performed in the presence of a non-biotinylated dsx pre-mRNA containing 6xGAA repeats. Selections were performed using dsx pre-mRNA with no enhancer sequence (lanes 3,6,9), with the 3xGAA enhancer (lanes 4,7,10), or with the 6xGAA enhancer (lanes 5,8,11,12) from splicing reactions containing "mock" depleted extract (CtrlΔ, lanes 2-5), U1 snRNP-depleted extract (U1Δ, lanes 6-8) or a U2 snRNP-depleted extract (U2Δ, lanes 9-11). The selection in lane 12 was performed from a splicing reaction containing an equal mixture of the U1 and U2 snRNP-depleted extracts. Lane 1 contains RNA recovered directly from nuclear extract, representing approximately 3% of the amount of extract used in each selection.
6B. A subpopulation of U2 snRNP associates with SRm160/300 in the absence of exogenous pre-mRNA. RNA recovered after immunoprecipitation with rAb-SRm160 (lane 3) or a corresponding pre-immune serum (lane 2) from HeLa nuclear extract (lane 1) was analyzed as in 6A. The amount of nuclear extract represented in lane 1 corresponds to approximately 5% of the amount used for each immunoprecipitation.

Figure 7. SRm160/300 interacts with the ESE-binding protein hTra2β.
Immunoprecipitates were collected from HeLa nuclear extract using mAb-B1C8 (lanes 4 and 5) and a control monoclonal antibody (B3; specific for the hyperphosphorylated large subunit of RNA polymerase II; lane 3), transferred to nitrocellulose and immunoblotted with an affinity purified anti-peptide antibody specific for hTra2β. Total nuclear extract separated in lanes 1 and 2 represents approximately 10% of the amount of extract used in each immunoprecipitation. Nuclear extract was pre-incubated in the presence (lanes 1, 3 and 4) or absence (lanes 2 and 5) of ribonuclease prior to immunoprecipitation. Size markers (in kDa) and the rabbit immunoglobulin (Ig) heavy chain, derived from rabbit anti-mouse antibody used to couple mAbs-B1C8 and B3 to protein A Sepharose, are indicated.

Figure 8. Model for the role of the SRm160/300 splicing coactivator in Exonic Splicing Enhancer (ESE) function.

Interactions involving the binding of hTra2β to a GAA-repeat ESE and U1 snRNP to the 5′ splice site are required to recruit SRm160/300 to the pre-mRNA. Since neither of these interactions alone are sufficient for SRm160/300 recruitment, it is proposed that critical interactions mediated by one or more SR proteins between SRm160/300 and U1 snRNP, and between SRm160/300 and the ESE, promote the formation of spliceosomes and splicing. These interactions simultaneously recruit U2 snRNP to the pre-mRNA, which also interacts with SRm160/300. It is also proposed that these interactions promote the pairing of specific pairs of exons during the regulation of splicing site selection.

Table 1.

Repetitive sequences in the SRm300 ORF[1]

Repeat Consensus	Location[2]
(PA)**RRGRSR**(SRTP)	439, 477, 506, 517, 527, 558, 589, 601, 612, 624, 635, 647, 658, 667, 1749
(SX)**SSPE**(PK)	692, 729, 768, 787, 862, 1167, 1370, 1409, 1431, 1448, 1528, 1547, 1567, 1621, 1640, 1678
SRSGSS(S)**P**	727, 746, 839, 857, 878, 917, 937, 957, 1388, 1407, 1526, 1545, 1656

[1]The SRm300 ORF sequence was analyzed for statistically significant repetitive sequences using the SAPS program (Brendel et al., 1992). Many repetitive sequences scored as significant, consisting of separated, tandem or overlapping repeats. Only consensus sequences of repeats occurring ten or more times are listed above; the corresponding sequences are boxed in Figure 1. Consensus sequences highlighted in bold type are present in all of the repeats listed whereas the sequences in parentheses are present in only a subset of these repeats. "X" denotes any amino acid residue.
[2]Represented as the first amino acid residue in the repeat consensus (as boxed in figure 1A).

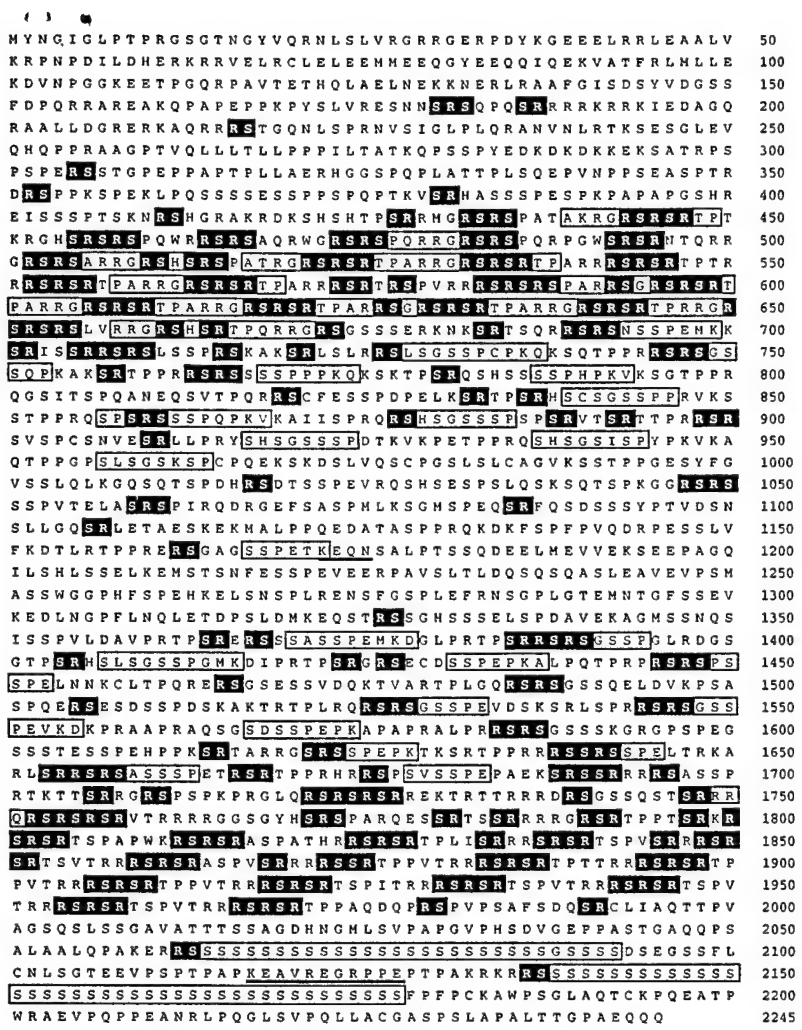


Figure 1A

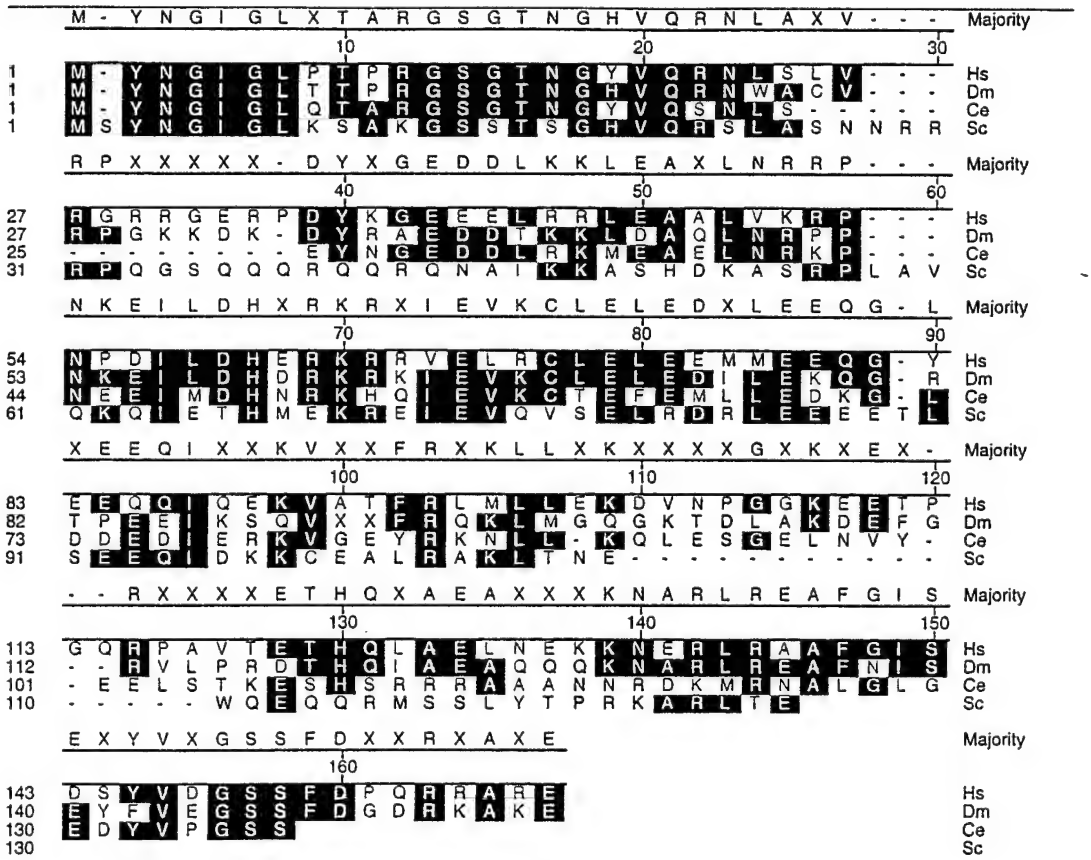


Figure 1B

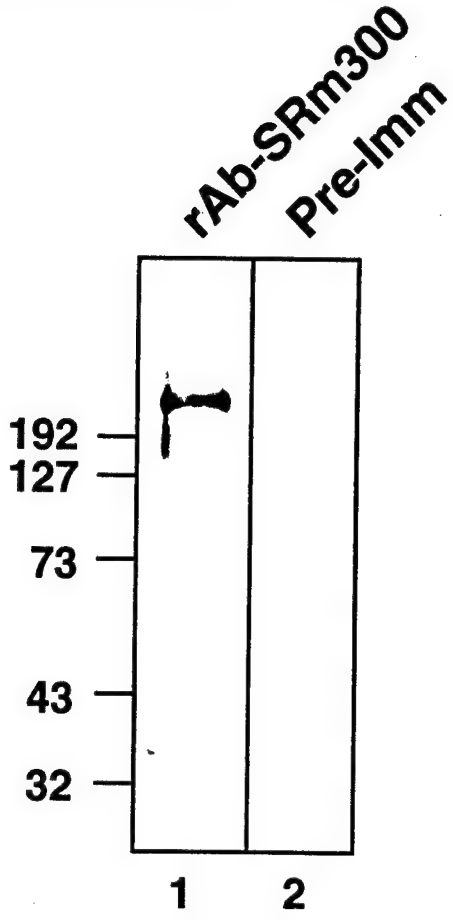


Figure 2A

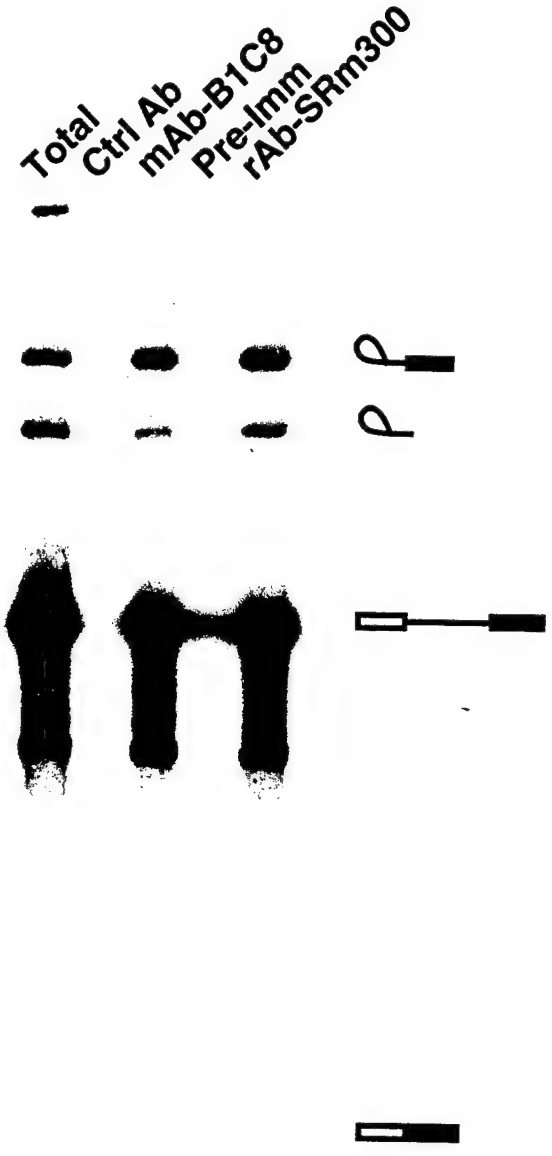


Figure 2B

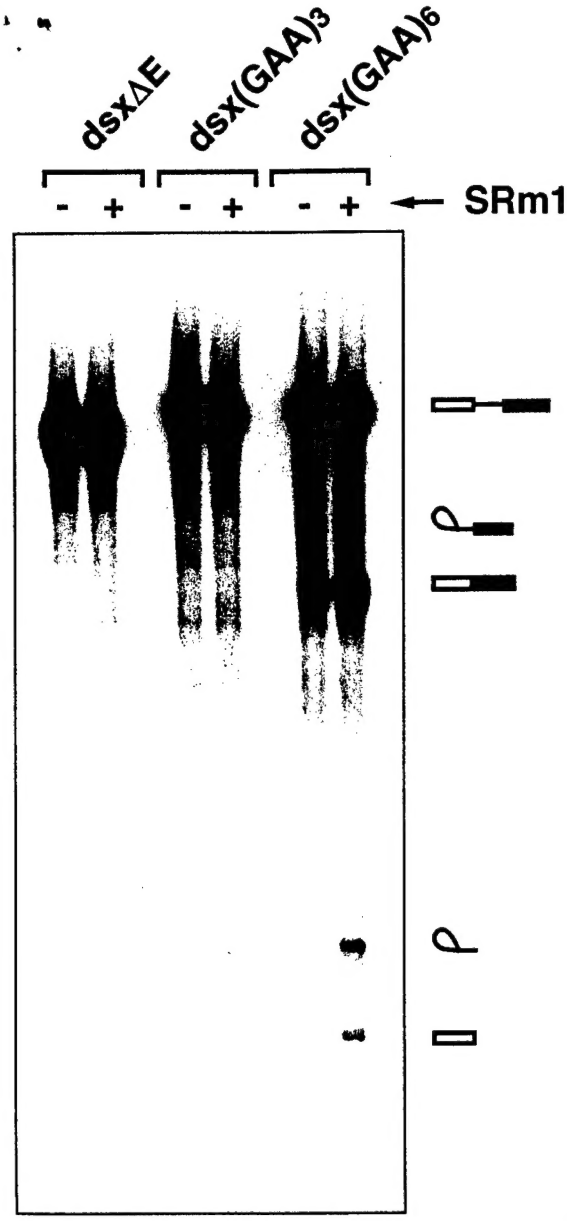


Figure 3

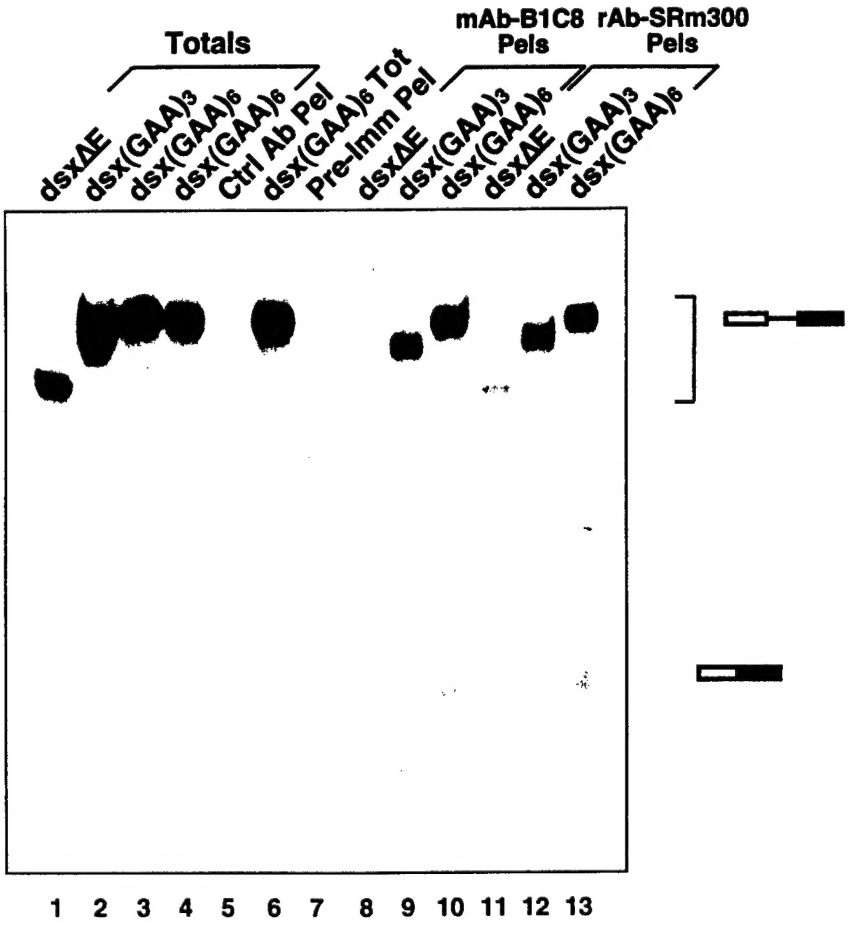


Figure 4

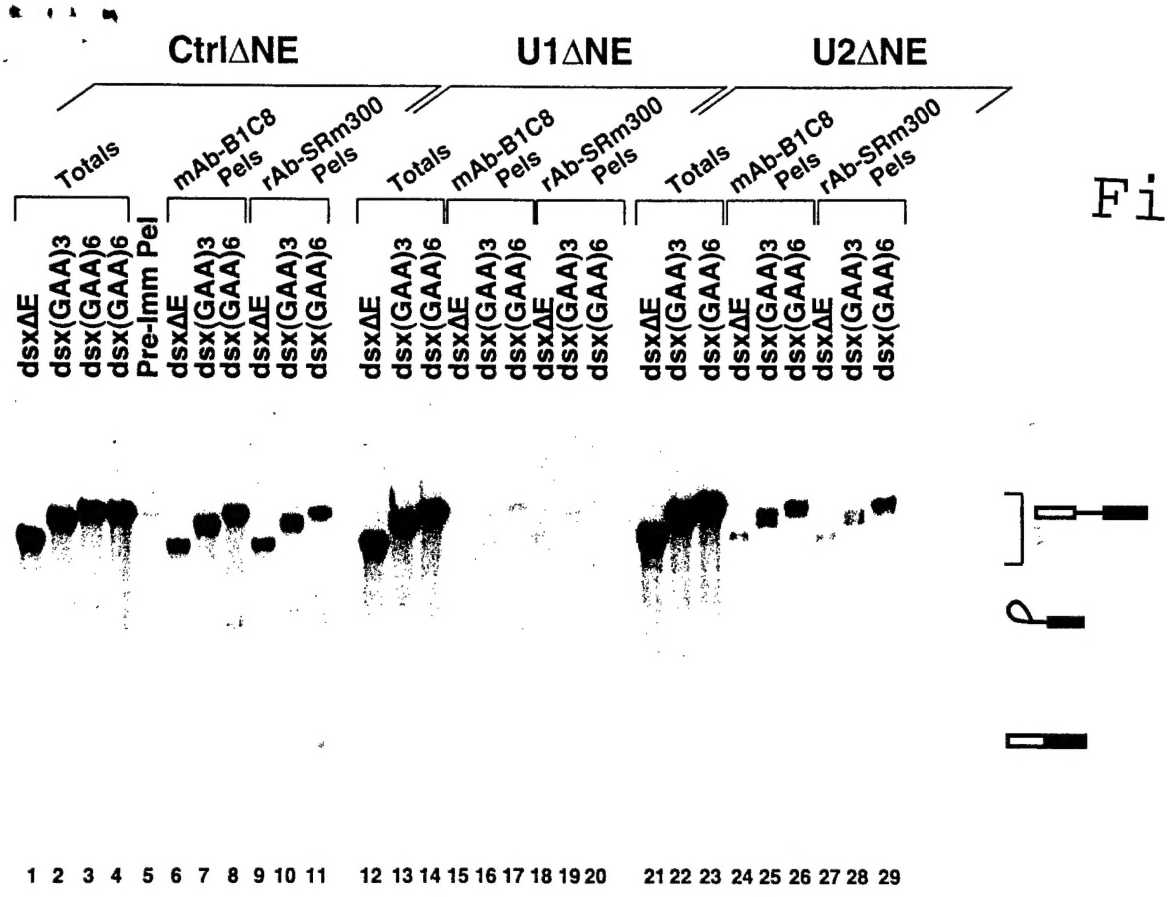


Figure 5

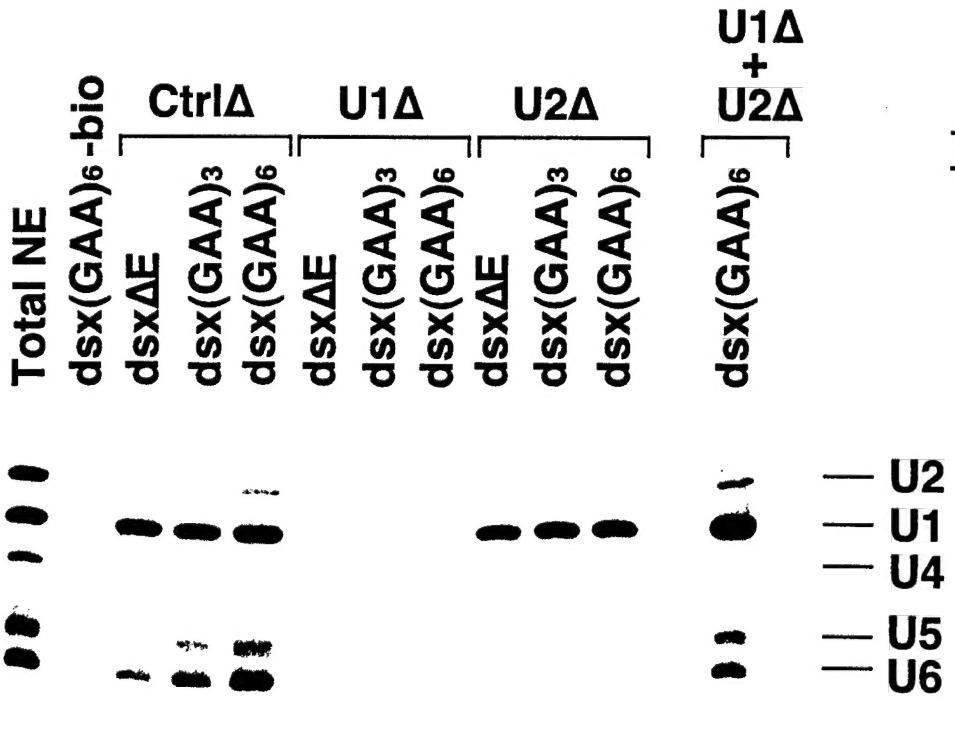


Figure 6A

Figure 6B

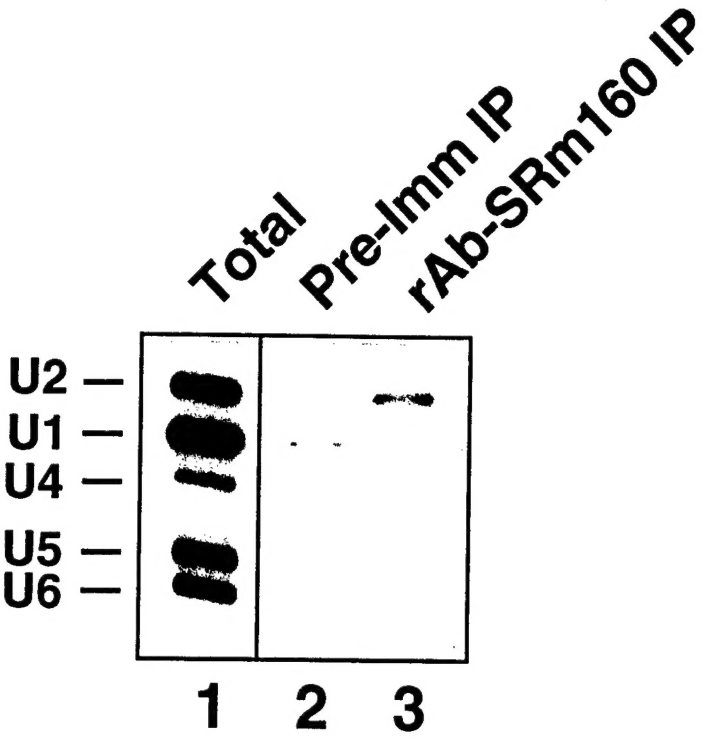
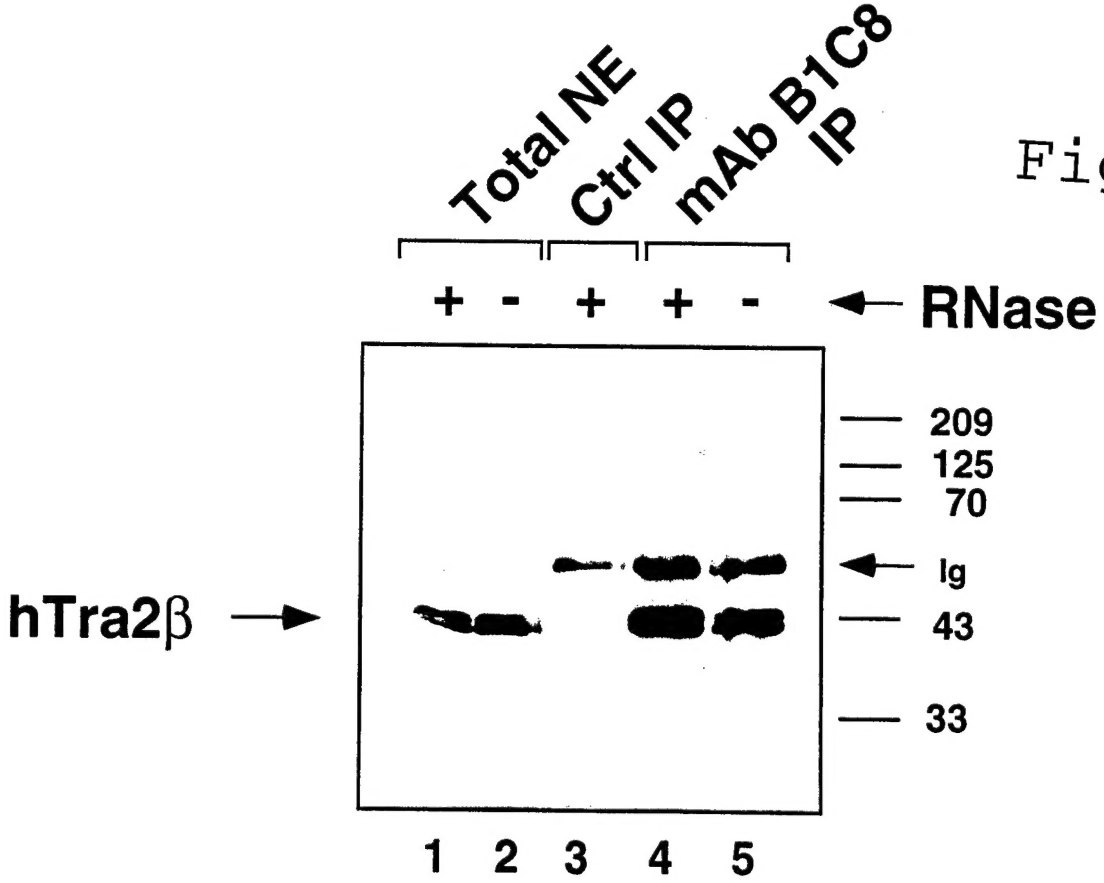


Figure 7



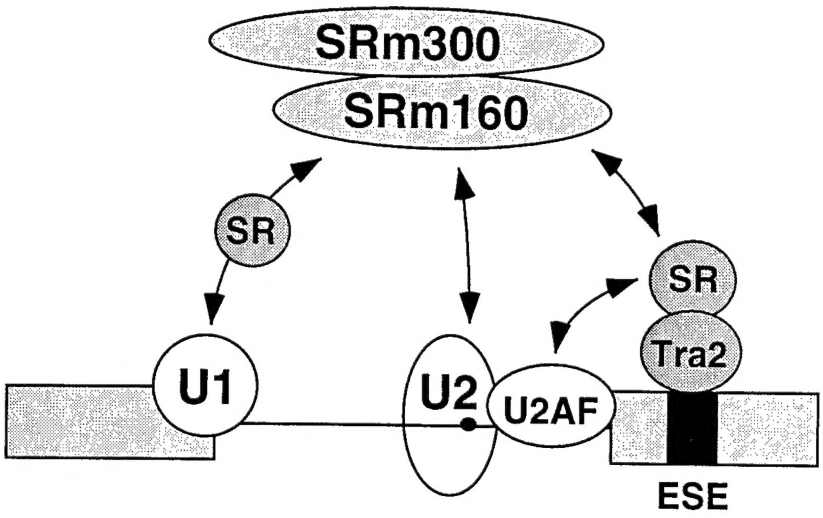


Figure 8